CLASSIFIERS BASED ON ARTIFICIAL INTELLIGENCE TECHNIQUES FOR THE DIAGNOSIS OF LUNG CANCER

VIJAY L. AGRAWAL[1] & S. V. DUDUL[2]

[1]Faculty Department of Electronics and Telecommunication, HVPM''s College of Engineering & Tech.,
Amravati, Maharashtra, India

[2]Department of Applied Electronics, Sant Gadge Baba Amravati University, Amravati, Maharashtra, India

ABSTRACT

In order to develop algorithm 39 CT scan images of patients have been considered consisting of Benign Tumor, Malignant Tumor and Normal Lung CT Scan iimage. With a view to extract features from the CT scan images after image processing, an algorithm is developed which proposes two-dimensional discrete cosine Transform domain coefficients in addition to Average, Standard Deviation, Entropy, Contrast, Correlation, Energy, Homogeneity. The suitability of classifiers based on Multilayer Perceptron (MLP) Neural Network is explored with the optimization of their respective parameters in view of reduction in time as well as space complexity. A separate Cross-Validation dataset is used for proper evaluation of the proposed classification algorithm with respect to important performance measures, such as MSE and classification accuracy. The Average Classification Accuracy of MLP Neural Network comprising of one hidden layers with 7 PE's organized in a typical topology is found to be superior (100 %) for Training . Finally, optimal algorithm has been developed on the basis of the best classifier performance.

KEYWORDS: Optimal Classifer, MLP, Computational Intelligence for Diagnosis of Lung Cancer, Diagnosis of Lung Cancer by Computational Intelligence Technique, Optimal Classifiers for Lung Cancer, Neural Network, Lung CT Images, Cross Validation for Lung Cancer

INTRODUCTION

Cancer is a petrifying disease, death-dealing disease. The sufferer alone can know the torment it causes.

There are many types of cancers. Lung cancer is one of the most common and deadly diseases in the world. The Incidence, Lung cancer is on Second Top and the Highest in death rate. It is a dreaded cancer disease for the human death.

These patients are not confirmed with cancer & treated wrongly in early stages due to lack of experts, clinical interpreters. The delay in detection, false diagnosis by experts, lack of experts in small towns, costly diagnosis are some of the reasons to these hapless victims for increase in death rate.

To mitigate their sufferings, an expert Lung cancer diagnosis Computation Intelligence system has been developed where experts could get second opinion for the confirmation of the disease in its early, curable stage.

In this paper optimal classifier based on Computational Intelligence techniques for the diagnosis of Lung Cancer has been developed.

After regrious training & retraining of the classifier, it is cross validated & tested on the basis of many performance matrix.

Use of the optimal classifier based on Computational Intelligence techniques results in more accurate and reliable diagnosis of lung cancer disease. To elevate the plight of poor patients, our optimal classifier will prove to be a major boon.

Our system will help in diagnosis of lung cancer disease in its early stage, consequently the survival rate of patient can be pro-longed with affordable low cost treatment, medication etc etc.

The proposed algorithm provides classification using classifiers based on Multi-layer Perceptron neural network approach and tested on the Lung CT scan images comprising of features extracted using 2D DCT domain co-efficient.

FEATURE EXTRACTION

Collected Lung CT images are in. jpg format. By using image processing & cropping the region of interest (ROI) the 128 features are extracted.
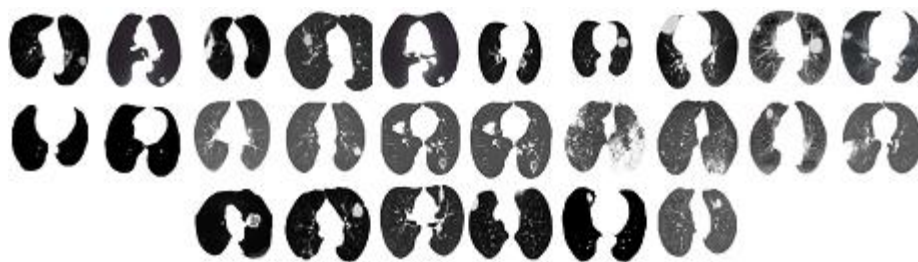

Figure 1: Few Samples of Input Processed Images of Lung

(Above lung images are of Benign, Malignant and Normal types)

Each Lung CT image is represented by a feature vector, F; which is comprised of 128 different parameters. The dataset contains 39 instances (exemplars) for three different classification. The classifier based on neural network is trained from the training dataset, where a feature vector is mapped on to a particular class or name of the Lung disease. The neural network learns from data (training exemplars) and the connection weights and biases are estimated as a result of this learning. After training of the neural network, its connection weights are frozen and latter; it is tested on a different dataset, which was never presented to the neural network. Here, this dataset is known as a cross-validation (CV) dataset. The performance of the classifier based on neural network is evaluated on the basis of some metrics, such as, MSE, NMSE, Classification Accuracy and Confusion Matrix. In this work, the prototype model of the classifier is developed with a view to discriminate between 3 different lung diseases. However, the proposed algorithm can be easily applied for classification of more than 3 lung diseases provided that one has enough computational resources. The feature vector, which is to be extracted from the separated ROI of Lung image, is as follows.

F = [DCT1, DCT 2, DCT 3, ..., DCT 128, Average, Standard Deviation, Entropy, Contrast, Correlation, Energy, Homogeneity, Shape];

Where DCT 1, DCT 2, DCT3, ..., DCT 128 denote the two-dimensional discrete Cosine transform domain coefficients.

EXPERIMENTAL SETUP

When working with large images, normal image processing techniques may sometimes break down, because the images can either be too large to load into memory, or else they can be loaded into memory but then be too large to process.

To avoid these problems, block-processing approach is used, where one can process large images incrementally:

reading, processing, and finally writing the results back to disk, one region at a time. In block-processing, an image, a block size, and a function handle are specified and then the input image is divided into blocks of the specified size. Later, all blocks are processed using the function handle one block at a time, and then the results are assembled into an output image. For lung images, block size of 16x16 is used for optimal results as compared to block size of 4x4 and 8x8.

An environment, accessible from MATLAB R2010b (Mathworks Inc., USA) is used to implement the algorithm that processes the input image resulting in 2D discrete WHT domain coefficients in addition to Average, Standard Deviation, Entropy, Contrast, Correlation, Energy, Homogeneity and Shape descriptor. Here class and shape descriptor are symbolic or qualitative, whereas all other parameters are numeric-valued or quantitative. The values obtained were exported to spreadsheet.

Neural Networks: Neuro Solutions (Neuro Dimensions, Inc. USA) 5.0 was used to implement various NN based classifiers on lung image which is represented by a Feature Vector containing 128 different elements.

MLP based classifier were explored and studied with respect to the performance measures.

Performance Measures

MSE (Mean Square Error):

The formula for the mean squared error is:

$$MSE = \frac{\sum_{j=0}^{P}\sum_{i=0}^{N}\left(d_{ij}-y_{ij}\right)^2}{NP} \tag{1}$$

Where P = number of output processing elements, N = number of exemplars in the data set, y_{ij} = neural network output for exemplar i at processing element j, d_{ij} = desired output for exemplar i at processing element j.

NMSE (Normalized Mean Square Error):

The normalized mean squared error is defined by the following formula:

$$NMSE = \frac{P\,N\,MSE}{\sum_{j=0}^{P}\dfrac{N\sum_{i=0}^{N}d_{ij}^{2}-\left(\sum_{i=0}^{N}d_{ij}\right)^{2}}{N}} \tag{2}$$

Where P = number of output processing elements, N = number of exemplars in the data set, MSE = mean square error, d_{ij} = desired output for exemplar i at processing element j.

Confusion Matrix

A confusion matrix is a simple methodology for displaying the classification results of a network. The confusion matrix is defined by labeling the desired classification on the rows and the predicted classifications on the columns. Since we want the predicted classification to be the same as the desired classification, the ideal situation is to have all the exemplars end up on the diagonal cells of the matrix (the diagonal that connects the upper-left corner to the lower right).

However, now, we already know the number of PEs in the first hidden layer.

It is observed from the following Table 1 and figure 2 during Training that for 7 PEs in the first hidden layer, the

average of Minimum MSE on the CV dataset is the least. Therefore, our MLP NN should have 7 PEs in the hidden layer.

Table 1

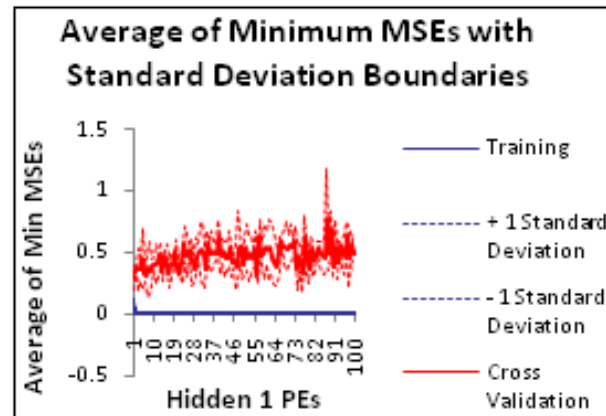| Best Networks	Training	Cross Validation
Hidden 1 PEs	31	7
Run #	3	2
Epoch #	1000	436
Minimum MSE	1.81527E-26	0.150729106
Final MSE	1.81527E-26	0.151806227

Figure 2

It is observed that MLP NN with one hidden layer with the 136-89-7-39 configuration yields the best results. From the above experimentation, selected parameters for designing optimum MLP NN classifier are given below:

No. of inputs = 136, No. of hidden layers=01,

No. of output PEs = 7, No. of epochs=1000,

For one hidden layer PEs and output layer PEs, for transfer function Tanh, Learning Rule Momentum NNhas been tested for training and testing the network.

RESULTS & CONCLUSIONS

For reliable clssification of lung images into three different types, classifier based on MLP NN have been developed and studied to get various variable parameters for optimum performance on Testing as well as cross-validation dataset.

The obtained Test Results are as shown belows:

Table 2

Output / Desired	Output(B)	Output (NL)	Output (M)
Output(B)	1	0	0
Output(NL)	0	0	0
Output(M)	1	1	1

Table 3

Performance	Output(B)	Output(NL)	Output(M)
MSE	0.138479267	0.261424548	0.186867629
NMSE	0.55391707	1.394264254	0.996627356

Table 3: Contd.,			
MAE	0.343045626	0.411481493	0.40802023
Min Abs Error	0.155631427	0.175869294	0.24379757
Max Abs Error	0.510060686	0.932950742	0.619114861
r	0.748759704	-0.887960798	0.431000767
Percent Correct	50	0	100

The proposed classifier is noticed as 100% on Tested data set .

ACKNOWLEDGEMENTS

REFERENCES

1. IEEE Transactions on Medical Imaging (April 1996, Vol. 15, No. 2) : Jyh-Shyan Lin

2. IEEE transaction on Medical Imaging(Dec.1998 , Vol. 17, No. 6): Manual G. Penedo

3. American Roentgen Ray Society (March 2002, AJR : 178): Yuichi Matsuki

4. CHEST journal (January 2003): Bethany B. Tan, Kevin R. Flherty

5. American Association for Artificial Intelligence(Flair-2003): C. F. Aliferis0

6. Application of a Neural Network to Improve Nodal Staging Accuracy with F-FDGPET in Non-Small Cell Lung cancer (2003 ; 44:1918-1926): Hubert Vesselle Radiology(Nov. 2005): Kwang Gi. Kim

7. The Journal of Nuclear Medicine (2006 ; 47:1075-1080): Yongkang Nie

8. Proceedings of the WSEAS International Conference on Signal, Speech and Image processing Lisbon, Portugal, (September 22-24, 2006): Kim Le

9. Medical Physics(March 2007,34(3)): Peng Wang IEEE/ICME: JIA TONG

10. Cancer Biomarkers, IOC Press(2007): Michael Phillips, Nasser Altorki

11. International Journal of Computer Applications,(2010,Vol.1-No.4): K. Balachandran, Dr. R. Anithha

12. International Journal of Bio-science and Bio-Technology(June 2010 , Vol. 2 , No. 2) : Rishi Pal, Pradeep Garg

13. IACIST International Journal of Engg. & technology (June 2010 , Vol. 2, No. 3) : M. Gomathi and Dr. P. Thangaraj

14. Digital Image Processing ,second edition, PHI publication: Rafael C. Gonzalez

15. Digital Signal Processing –Principles, Algorithms, and Applications, Fourth edition ,Pearson Education: John G. Proakis, Dimitris G. manolakis

16. New Atlas of Human Anatomy, Lustre Press Pvt. Ltd. : Thomas McCracken, General Editor

17. http://www.wrongdiagnosis.com/c/cancer/intro.htm?ktrack=kcplink